2011-10-12

Bayesian Networks for Real Caller ID, Project Final Report

Volpano, Dennis

http://hdl.handle.net/10945/37267

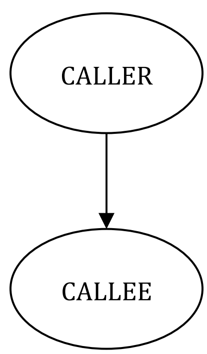
Bayesian Networks for Real Caller ID

Project Final Report[1]

Dennis Volpano

Speaker recognition is just one input to deciding caller ID. It relies exclusively upon a voice sample taken from a caller. There is more information about a caller and their phone that can be leveraged to identify them, or narrow the possibilities at least, independently of voice. It includes most frequently called parties as conveyed by a contact list or speed-dialing selections, caller location, calling habits (time of day and frequency) and phone battery level. All contribute toward a passive caller signature that can be used to track the most recent phone used by a person (e.g. in a re-direction service that allows users to reach the person using one phone number regardless of which phone the caller is currently using) or to distinguish among multiple users sharing a phone in order to provide a caller ID before the caller says anything. Speaker recognition at best can say something about the history of a phone's usage which may narrow caller ID possibilities. Until the caller speaks however another signature is needed.

We illustrate a call-behavior approach to caller ID by building a simple Bayesian network. The network will, when executed, compute probabilities of interest such as the distribution among possible callers knowing only a dialed number or knowing a dialed number and the phone from which it was dialed. We begin with a simple Bayesian network that accounts only for a user's frequently-dialed numbers as determined by the contact list on their phone:



As a causal network, it captures our interest in knowing the effect a caller has on who is called. The Asterisk call server provides for each call, one of three states (2002, 2003, 2004) of the variable CALLEE, and one of three states (1002, 1003, 1004) of the variable CALLER-EXT. The former numbers represent public phone numbers that would be issued to users, one per user, and the latter numbers are phone extensions used by Asterisk and SIP clients. Users do not dial phone extensions. They dial only numbers in the 2000 series.

[1] We thank the Marine Corps Warfighting Lab for supporting this project.

Suppose for a given call, Asterisk has provided state c of CALLEE (ignore the state of CALLER-EXT for a moment). We introduce a variable CALLER with the same states as those of CALLEE. We want for each state v of CALLER, P(CALLER=v | CALLEE=c):

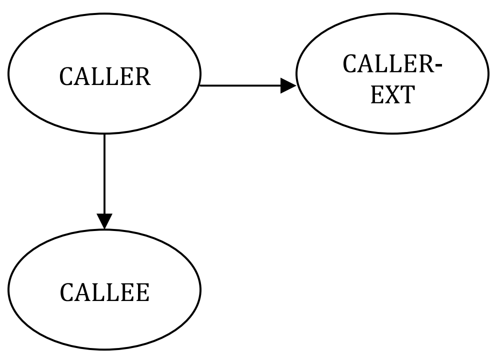
$$
\begin{aligned}
& P(\text{CALLER=v} \mid \text{CALLEE=c}) & \\
= \; & P(\text{CALLER=v, CALLEE=c}) \,/\, P(\text{CALLEE=c}) & \text{conditional probability formula} \\
= \; & P(\text{CALLEE=c} \mid \text{CALLER=v})P(\text{CALLER=v}) \,/\, P(\text{CALLEE=c}) & \text{joint probability definition} \\
= \; & \frac{P(\text{CALLEE=c} \mid \text{CALLER=v})P(\text{CALLER=v})}{\sum_{\text{CALLER} \in \{2002,2003,2004\}} P(\text{CALLEE=c, CALLER})} & \text{Caller marginalization} \\
= \; & \frac{P(\text{CALLEE=c} \mid \text{CALLER=v})P(\text{CALLER=v})}{\sum_{i = \{2002,2003,2004\}} P(\text{CALLEE=c} \mid \text{CALLER=i})P(\text{CALLER=i})} &
\end{aligned}
$$

If P(CALLEE=c | CALLER=v) = 1/d where d is the number of entries in the contact list of v if c is in v's contact list and zero otherwise, and P(CALLER=v) = 1/|CALLER| (this is oversimplified as the distribution is not normally uniform in practice) then the denominator above becomes just

$$1/n \sum_{i=1,\ldots,k} 1/d_i$$

where k is the number of contact lists containing c, $d_1,\ldots,d_k$ are the numbers of entries respectively in the k lists containing c and n is the total number of contact lists (we assume each caller has a contact list so the total number of contact lists is the total number of callers, or n=3).

Next we leverage the variable CALLER-EXT. Our Bayesian network becomes:



Again as a causal network, we are interested in knowing the effect a caller has on the extension used to place a call. Suppose Asterisk provides state ce of CALLER-EXT for a given call and state c of CALLEE. Then we want P(CALLER=v | CALLEE=c, CALLER-EXT=ce):

P(CALLER=v | CALLEE=c, CALLER-EXT=ce)

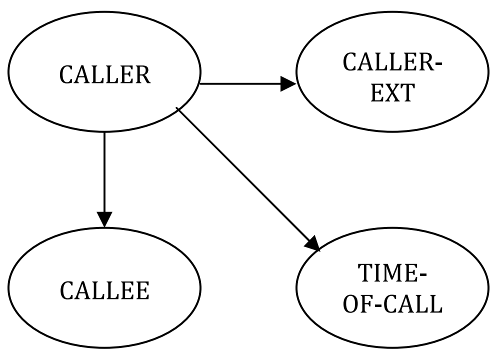
$$= \frac{P(\text{CALLER=v, CALLEE=c, CALLER-EXT=ce})}{P(\text{CALLEE=c, CALLER-EXT=ce})} \qquad \text{conditional probability formula}$$

$$= \frac{P(\text{CALLEE=c} \mid \text{CALLER=v})P(\text{CALLER-EXT=ce} \mid \text{CALLER=v})P(\text{CALLER=v})}{\sum_{\text{CALLER} \in \{2002,2003,2004\}} P(\text{CALLEE=c, CALLER, CALLER-EXT=ce})}$$

$$= \frac{P(\text{CALLEE=c} \mid \text{CALLER=v})P(\text{CALLER-EXT=ce} \mid \text{CALLER=v})P(\text{CALLER=v})}{\sum_{i \in \{2002,2003,2004\}} P(\text{CALLEE=c} \mid \text{CALLER=i})P(\text{CALLER-EXT=ce} \mid \text{CALLER=i})P(\text{CALLER=i})}$$

We expect P(CALLER-EXT=ce | CALLER=v) to account for the physical proximity of v to extension ce, the battery level of v's own phone and also that of ce. A higher conditional probability would arise when v has recently been in close proximity to ce, ce is adequately charged and if ce is not v's own phone then v's phone is not charged or has been lost or stolen.

It's straightforward to extend the Bayesian network with more information provided by Asterisk. One obvious extension is a time-of-call variable with some finite set of states corresponding to a chosen discrete resolution of time:

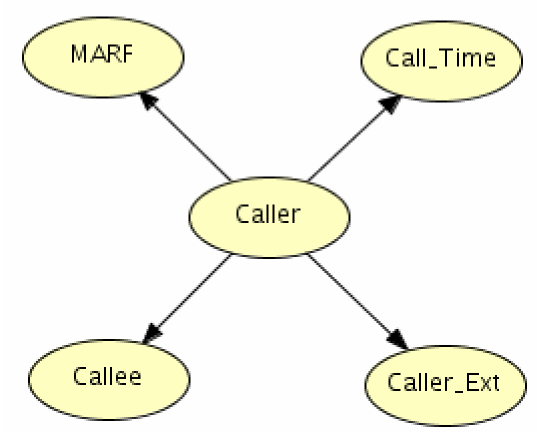


Now the network captures the effect a caller has on the time a call is placed so like before, we see an edge from CALLER to TIME-OF-CALL.

A final extension we can make to the network comes if we have a voice sample of the caller when the call is placed, for instance, with voice-activated calling. We add a variable for MARF with three states: 2002, 2003 and 2004. MARF (Modular Audio Recognition Framework [2]) is open-source speaker recognition technology. Which state is true is determined by MARF's first choice after it attempts to recognize the caller's voice. We populate the table for P(MARF=x | CALLER=y) where x and y range separately over 2002, 2003 and 2004 according to an experiment with MARF. MARF has training and testing samples from each of the users identified by these 2000-series extensions. For example, suppose we have 5 testing samples for user 2002. On two of them, MARF correctly identifies the user. Of the remaining 3, two are identified as 2003 and one as

2004. Then P(MARF=2002 | CALLER=2002) = 2/5, P(MARF=2003 | CALLER=2002) = 2/5 and P(MARF=2004 | CALLER=2002) = 1/5.[2]

The Bayesian network now becomes



Around the perimeter of the network are evidence variables and they are independent of each other. Hence the network is a Naïve Bayesian network.

We want P(CALLER=v | CALLEE=c, CALLER-EXT=ce, CALL-TIME=t, MARF=u) after a call for each user v where c, ce and t are evidence provided by the call server and u is evidence provided by MARF (MARF's first choice in recognizing the caller's voice sample). We have

$$P(\text{CALLER=v} \mid \text{CALLEE=c, CALLER-EXT=ce, CALL-TIME=t, MARF=u})$$

$$= \frac{P(\text{CALLER=v, CALLEE=c, CALLER-EXT=ce, CALL-TIME=t, MARF=u})}{P(\text{CALLEE=c, CALLER-EXT=ce, CALL-TIME=t, MARF=u})}$$

A sample run of the above network in Hugin 7.3 is shown in Figure 1. It shows the results of caller ID (variable Caller) for a call placed from extension 1004 to user 2003 in the afternoon where the voice of the caller is analyzed by MARF and determined by MARF to be that of user 2004. There's roughly a 59% chance the caller is user 2004 compared to a 41% chance it is user 2002 (it is zero for user 2003 since no user calls himself).

Built into this Bayesian network is a call structure where users 2002 and 2004 are in proximity to one another but each calls the other infrequently (e.g. two squad leaders who only communicate to their platoon leader). User 2003 calls 2002 and 2004 with equal probability (user 2003 is the platoon leader).

---

[2] If MARF were to say unknown in response to n of the 5 samples from 2002 then we would distribute mass n/5 across all 3 users by adding n/(3*5) to each of the probabilities conditioned on CALLER=2002. It is sound to do so since caller identity is not determined by the conditional probability alone for a user but rather by how much it differs from other users as well.

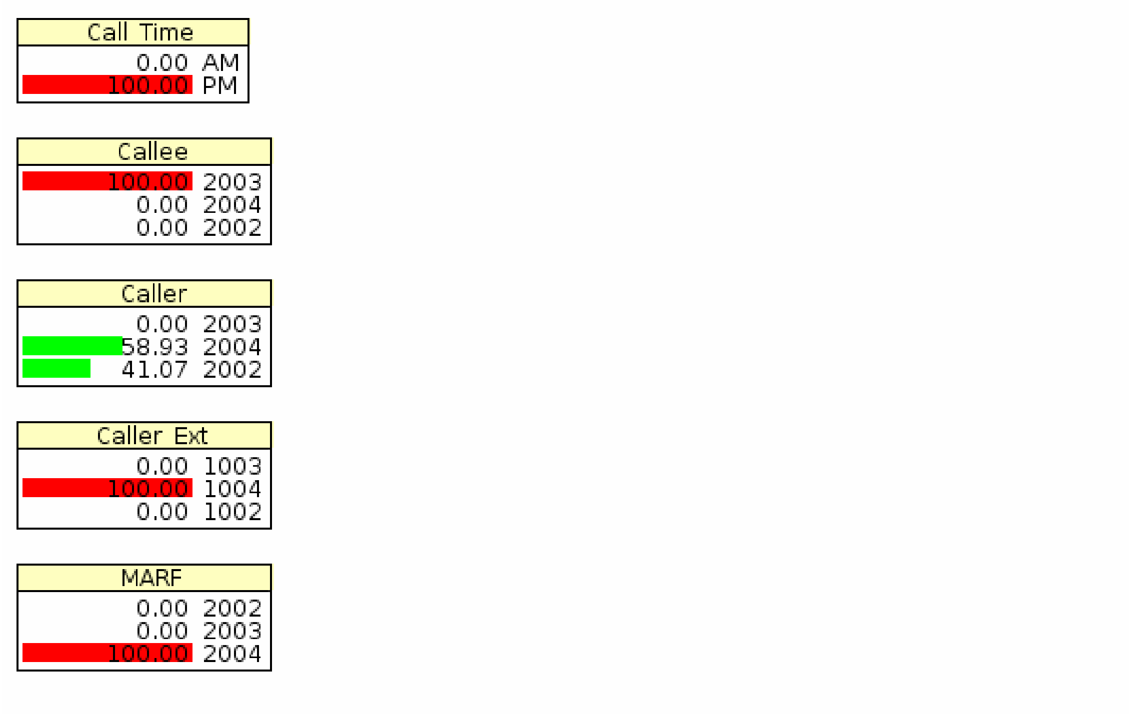


Figure 1. Hugin 7.3 display of Bayes network execution

As an instance of a Naive Bayes network, the caller could be identified by a much simpler argmax computation. Let F be the feature vector that instantiates the four evidence variables after a call (suppose these variables are renamed to $f_1,\ldots,f_4$). By Bayes rule,

$$P(CALLER=v \mid F) = P(F \mid CALLER=v)P(CALLER=v)/P(F).$$

We want the caller who maximizes v. In other words, we want,

$$\underset{v}{\operatorname{argmax}}\ P(F \mid CALLER=v)P(CALLER=v)/P(F)$$

or simply argmax P(F | CALLER=v)P(CALLER=v) since P(F) is independent of v. Thus we have

$$\underset{v}{\operatorname{argmax}}\ P(F \mid CALLER=v)P(CALLER=v)$$
$$= \underset{v}{\operatorname{argmax}}\ P(CALLER=v) \prod_{i=1}^{4} P(f_i \mid CALLER=v)$$

Training

Where do the conditional probabilities in the Bayesian network come from? In general, the probability tables would be fed by observing a statistically significantly number of calls among users in a population. Military application of cellular phones appears to reduce this need somewhat. Mission requirements may dictate a disciplined calling pattern among users that can make it easier to build probability tables. For instance, we can exploit the communication structure between two squad leaders and their platoon leader. A mission might require that squad leaders do not call each other but only their platoon leader.

Updates to the probability tables must also be made to reflect changing conditions in the world. Suppose users 2002 and 2004 rotate running reconnaissance patrols. While one is out the other stays behind to pull security, protecting the perimeter of user 2003. Upon returning from a patrol, user 2004 discovers that user 2002 has temporarily damaged his phone (device 1002) by accidentally dropping it in the water. So user 2002 takes user 2004's phone (device 1004) on patrol leaving 2004 without a useable phone. Can our Bayesian network show us that calls placed by user 2002 from device 1004 are actually from user 2002? That will depend on how quickly the network can be updated. We would reduce P(CALLER=2004) to reflect user 2004's inability to place a call and increase P(CALLER-EXT=1004 | CALLER=2002). When and how such updates occur is the subject of further work.

Variable MARF might be updated over time if there is a secure way to provide feedback to the call server. For example, extending the example above, if after a call from user 2002, the callee can confirm the caller was indeed 2002 and after analyzing the caller's voice, MARF determined it is from 2002 then P(MARF=2002 | CALLER=2002) = 3/6 (5 original testing samples of which MARF only identified 2 as 2002), P(MARF=2003 | CALLER=2002) = 2/6 and P(MARF=2004 | CALLER=2002) = 1/6. If instead MARF says the caller is 2003 then P(MARF=2002 | CALLER=2002) becomes 2/6 while P(MARF=2003 | CALLER=2002) = 3/6 and P(MARF=2004 | CALLER=2002) = 1/6.

References

[1] Jensen, Finn V. An Introduction to Bayesian Networks, Springer-Verlag New York, 1996.

[2] The Modular Audio Recognition Framework, http://marf.sourceforge.net.